
Mobile multimedia object management for efficient content search and exchange

C.H.C. Leung*

Victoria University, Melbourne, Australia

E-mail: Clement.Leung@vu.edu.au

*Corresponding author

T.L.Y. Leung

University of London, London, UK

E-mail: timothyleung@gmail.com

Abstract: In Mobile Multimedia Object (M³O) retrieval, high precision is required, as there is limited scope for browsing or relevance feedback due to limitations in both time and display size. In M³O, three kinds of data will need to be managed:

- raw content data
- metadata for assisting non-content based searches
- structured content description data to facilitate fast content based searches.

These three kinds of data have to be processed and integrated to support efficient M³O management and retrieval. This paper examines methods, both automatic and semi-automatic, for capturing data that will support accurate M³O data retrieval. A data architecture and organisation structure of mobile multimedia data is presented.

Keywords: content-based search; information exchange; mobile multimedia object; relevance feedback; metadata; MPEG-7 structured annotation description.

Reference to this paper should be made as follows: Leung, C.H.C. and Leung, T.L.Y (2008) 'Mobile multimedia object management for efficient content search and exchange', *Int. J. Virtual Technology and Multimedia*, Vol. 1, No. 1, pp.61–74.

Biographical notes: Clement H.C. Leung holds the Foundation Chair in Computer Science at Victoria University, Australia. Previously, he held an Established Chair in Computer Science at the University of London, UK. His principal research interests are in the areas of Multimedia and Visual Information Systems, and High Performance Database Systems. His services to the research community include serving as Program Chair, Program Co-Chair, Keynote Speaker, Panel Expert, on the editorial board of several journals as well as on the Program Committee and Steering Committee of major international conferences.

Timothy L.Y. Leung holds a first class Honours Degree in Computer Science from University College London, University of London. He is a consultant specialising in corporate multimedia deployment and integration, and financial business analytics. His research interests include personal multimedia

repository management, music query processing, algorithmic trading, and pattern discovery in multi-paradigm financial data mining.

1 Introduction

With the increase in both the storage capacity of mobile devices and the explosive growth in their sale and usage, the task of managing the variety of mobile multimedia objects has become important as well as complex (Banavar et al., 2000; Graham et al., 2002; Platt et al., 2002; Poon and Chen, 2004; Rodden, 2003; Smith et al., 1999). Even the smallest devices can increasingly support the use of storage cards whose capacity is fast approaching $\sim 10^{10}$ bytes, while the larger devices have a storage capacity close to $\sim 10^{12}$ bytes. With the rapidly expanding functionality of mobile devices, and the convergence of different device types (e.g., PDAs, phones, MP3 players, cameras, laptop computers) together with the ease and variety of modes by which they can communicate in a wireless fashion, multimedia objects are propagated and created at an unprecedented pace. A robust indexing mechanism is necessary to enable the retrieval, insertion, and deletion of accumulated data that will efficiently interoperate in a mobile environment (Gemmell and Lueder, 2003; Rodden, 2003). Although there are some existing methods for conventional data management, these do not seem adequate for managing large databases of multiple types of media data. Such a mechanism will require a dynamic and efficiently adaptable index for optimal performance, as well as incorporate the ability to support efficient content-based search and information exchange among mobile devices.

In mobile multimedia object (M³O) retrieval, high precision is required, as there is restricted scope for browsing or relevance feedback due to limitations in both time and display size. In M³O, three kinds of data will need to be managed:

- raw content data
- metadata for assisting non-content-based searches
- structured content description data to facilitate fast content-based searches.

For example, a given photograph may represent the raw content data; the GPS coordinates of where the photograph is taken would represent the metadata; while a standard MPEG-7 Structured Annotation Description (indicating particular visual objects and their relationships within the photograph) would represent the structured content description data. These three kinds of data have to be processed and integrated to support efficient M³O management and retrieval. Although some degree of background (pre-)processing may be carried out by making use of idle CPU cycles when the device is not in use, the efficient capture of the bulk of metadata and particularly content description data tends to be time consuming and expensive.

We shall examine methods, both automatic and semi-automatic, for capturing M³O data, and a data architecture supporting flexible content search is described. The data objects that the M³O management system is able to manage include:

- voice memos
- wave files
- SMS and MMS
- hand-drawn documents and text files
- MIDI and MP3 files
- cartoons and vector/raster graphics
- photographs
- maps
- video clips.

The M³O data model is designed with three kinds of users in mind:

- the enthusiastic hobbyist who predominantly uses mobile devices to store and playback music, still images and video files
- the mobile business executive or m-commerce analyst who will need multimedia business data at their fingertips, anytime, anywhere
- the hybrid user who combines a mixture of personal with business use.

The indexing and data management architecture will be geared to a high level of interoperability where object exchange among mobile devices will automatically trigger the transfer of self-descriptive data to enable efficient content search of the newly transmitted multimedia object in the receiving device.

2 Types of mobile multimedia data objects and their characteristics

An efficient mobile multimedia object management should be able to store, manage, and retrieve the following types of objects.

2.1 Audio-based objects

There are several types of common audio-based objects that need to be managed and retrieved. They include voice memo, wave, MIDI and MP3 song objects. Voice memos tend to be relatively short in duration and usually are authored by the owner of the mobile device, which may be stored as wave objects. In addition, there are wave objects other than voice memos that will also need to be stored. Usually associated with the voice memos and wave objects are title, timestamp, and duration, which may be used to identify a particular object. Parsing of wave-based audio content can be done by signal processing and spectrum analysis. Although it is unlikely that reliable automatic content extraction is possible due to the multiplexing of different sources of sound in real-life scenarios, some simple sound patterns can be detected using model-based approaches if the background noise (e.g., using noise gates) can be minimised. Conversion of voice memos to text by means of voice recognition software is becoming more common, so that searching these by content will be reduced to the much simpler problem of text pattern matching.

MIDI and MP3 objects are usually obtained from sources external to the mobile device. MIDI objects are much more structured than MP3 objects, and depending on the type of MIDI files (Type 0, Type 1, or Type 2 SMF), the amount of information will differ. MIDI objects have the advantage that they include a significant amount of content information – such as key, time signature, tempo, instrumental voices, note sequence – which may be extracted automatically and they can, either in isolation or in combination, often be used as search arguments for the sophisticated user. In the cases of Type 1 and Type 2 MIDI files, the chord sequence and harmony characteristics may also be automatically extracted. Unlike MIDI objects, MP3 songs are mostly searched on metadata such as artist, composer, genre, duration, which also apply to MIDI objects. Searching by content for MP3 files is generally not possible unless some pre-processing or manual input processes are deployed. Pre-processing here will involve the use of statistical and signal processing techniques such as note segmentation and pitch tracking (Ghias et al., 2002); such pre-processing is highly processor intensive and is best performed on more powerful desktop machines. Query models may include ‘query by humming’ or ‘query by playing’ for searching by precise melody or the melodic contour (which focuses on the approximate shape of the melody rather than the precise notes); ‘query by playing’ using the input mechanism in the mobile device will be generally more feasible since ‘query by humming’ will involve rather formidable signal processing algorithms to convert the query into discrete musical notes (Leung, 2007).

2.2 Image-based objects

Image-based objects can take the form of photographs or MMS. In the case of photographs, they can mainly be identified by metadata, such as title, originator and time. The location and related data may be automatically captured also if the mobile device has a GPS module (Bharat and Minakakis, 2003; Toyama et al., 2003). However, the semantic visual contents of photographs (e.g., a boy riding on an elephant) cannot be automatically captured with the current image processing technology (Leung and Sutanto, 1999; Over et al., 2004; Viola and Jones, 2001; Wilhelm et al., 2004). Nevertheless, supporting mechanisms, such as the MPEG-7 Structured Annotation Description Scheme, are available if searches need to be carried out on the semantic content of a photograph (Azzam et al., 2004a). The MPEG-7 Structured Annotation Description Scheme was found to be very useful in semantic content identification, whose basic framework consists of objects, subjects, indirect objects, actions, qualifiers and is based on the Ternary Fact Model (Leung and Sutanto, 1999; Zheng and Leung, 1996). Less semantically rich attributes, however, may be obtained from automatic algorithms but these are mainly confined to colour, texture, and shape, which are generally not found to be helpful to human perception and retrieval of visual objects.

In the case of MMS, there is usually accompanying text, so that carrying out search based on the text will be much more manageable than that based on the image or graphics. In any case, searching MMS by content may be reduced to either graphics-based search, image-based search or text-based search.

Image compression techniques are useful not only for transmission and size reduction, but can also be used for indexing and identification. Owing to the imprecise nature of queries, some degree of browsing through miniature icons or compressed images would seem to be useful operations in facilitating image retrieval.

2.3 Video-based objects

The most common video-based objects are video clips, which would typically be searched based on the usual metadata. For example, internet video search currently uses filename and surrounding text (Stevenson and Leung, 2005). As for content-based search, motion vectors, salient video stills, annotations and video partitioning are some of the methods employed for their indexing and retrieval. The spatio-temporal aspect of objects between frames is an important property that may need to be indexed. As in the case of photographs, automatic processing and extraction of semantic content is still rather limited. In addition, content search based on the dialogue track may be exploited. This allows the voice to be automatically converted to text using speech recognition software whereby text pattern matching techniques may be used.

Some of the image retrieval techniques are often useful in video-based systems. Representative frames may be identified and isolated in a video clip, and these can then be regarded as still images and processed using image-based techniques and algorithms.

2.4 Graphics-based objects

Objects in this category may include cartoons and graphics, hand-drawn documents and maps. Like most objects, metadata-based search may include retrieval based on originator, date, title and size. Line arts, synthetic images and computer animation may be described by their geometrical properties. VRML modelling and the Synthetic-Natural Hybrid Coding (SNHC) scheme in MPEG-4 for creating models of inanimate objects may be useful in the representation of synthetically generated contents.

The line, polygon, surface patches, Bezier curves, and other mathematical entities can be indexed numerically, and this form of content can be easily manipulated and computed. The motion of objects in computer animation can be traced and the sequences of events can be indexed. In some situations, it may be more efficient to index the characteristics of the underlying graphics generation algorithms, since fast searching and identification may be done using text-based pattern-matching procedures. Content-based retrieval based on features such as colour, texture and shape for this category of objects will be more natural than that for natural objects, so that automatic feature extraction mechanisms (e.g., colour histogram) may be used more effectively.

As for maps, mechanisms for content-based retrieval are already well developed so that basic search using place name and location can be carried out efficiently. Depending on the application, more sophisticated techniques based on Geographic Information Systems (GIS) procedures can also be usefully incorporated.

2.5 Text-based objects

Text-based objects are the most widely used in traditional systems and come in a variety of forms, including SMS, free-text documents, and structured records. Of the different types of objects, searching text-based objects by content will be the fastest and most efficient, as text-based search techniques are the most mature and well developed. Consequently, as indicated earlier, it is always advantageous to convert other multimedia data objects (wholly or partially) to text to facilitate content identification.

Keywords and structured text can be indexed using established automatic indexing algorithms (Baeza-Yates and Ribeiro-Neto, 1999). Relational and object-oriented

databases are the basic models for representing structured records. Although object-oriented databases offer greater scope for accommodating multimedia objects, they do not in general provide tools for their manipulation and identification. Free text such as the annotation of images and stratification of videos can be indexed using text-oriented databases.

3 Data properties for different types of mobile multimedia objects

In M³O management, apart from the raw multimedia data, metadata as well as content data to facilitate fast object identification and search will need to be managed. Table 1 shows these two forms of data for different types of multimedia objects.

Table 1 Data properties for different mobile multimedia objects

<i>Object type</i>	<i>Metadata</i>	<i>Content data</i>
Photograph	Title, date/time, originator, location	Image objects, colours, textures, shapes, lines, MPEG-7 descriptions
MIDI	Title, artist, composer, duration, genre	Key, time signature, note sequence, tempo, chord sequence, instruments
MP3	Title, artist, lyricist, composer, duration, genre	Similar to MIDI, plus possible lyrics and/or keywords
Wave	File name, originator, size	Similar to MIDI, plus possible lyrics and/or keywords
SMS	Date/time, originator	Keywords
Text document	Title, originator, size, date/time	Keywords
Voice Memo	Title, date/time, originator, location	Keywords
MMS	Title, date/time, originator, location	Image visual objects, keywords
Video clips	Title, date/time, originator, location, Duration	Visual objects, persons and characters, MPEG-7 descriptions, keywords
Maps	Region name, coordinates	Place names, salient features (e.g., cultural, physical etc.)
Graphics, hand-drawn/ hand-written documents	Title, date/time, originator	Colours, shapes, lines, keywords

The extraction and capture of these data can take three forms:

- fully automatic
- semi-automatic
- manual.

Fully automatic direct extraction tends to apply more to metadata than content data. Such data are often somewhat limited in scope and used to support the more basic management and search functions. For example, the duration of a song, and the date/time of an MMS may be obtained automatically. Semi-automatic content extraction can be defined as requiring user-interaction at some point in order to input the metadata, as opposed to a fully automated process. These metadata are then transmitted with the data file and everyone who utilises the data file will also be able to access the accompanying metadata. Two problems immediately arise from this situation.

First, the metadata must be ideally computed at the creation of the data file, to ensure consistency across all identical files. In other words, suppose at a particular time, the file is created and propagated, and after a while, a certain user decides to input the metadata for the file he or she possesses. In such a situation, the metadata will only be included in subsequent copies of the file that originate from that latter user. Essentially, for unification to occur, metadata should be produced at the same time as the file.

Second, users are prone to making mistakes. The users that enter the metadata may be biased, or in some circumstances, even malicious (e.g., having consistent metadata in Trojans will increase their authenticity, hence raising the susceptibility of the user on the receiving end). Currently, there is no check to ensure that the metadata correspond to the data file, and the sheer volume of data, both being produced and currently in existence, does not help the cause. One temporary solution to remedy the situation is for a central database to exist. For example, with MP3 files, though conversion (or ripping) from CDs to MP3s may occur at different times, with different users, the metadata is able to be unified and synchronised via the internet through the Compact Disc DataBase (CDDB). The CDDB works on the principle of identifying individual CDs by hashing their table of contents. The hashed output is then compared with data records in the existing database and loads the corresponding metadata for that disc. The metadata in MP3 files is often in ID3v1 or ID3v2 format. This format acts as a container and has fixed bytes pertaining to various information relating to the MP3 files, including song title, artist, album, genre and even lyrics. Although this may aid the cause for unification, the problem still remains: a user must first enter the metadata onto the CDDB.

The more flexible human-oriented retrieval criteria tend to require manual input, as they encroach on the limits of current technology. For example, there are no known algorithms for automatic recognition of visual objects in images, and this problem is unlikely to be resolved in the near future. Thus, manual processing of the image is required. Although semantic modelling (Zheng and Leung, 1996), automatic index expansion (Azzam et al., 2004b) and the establishment of the MPEG-7 Structured Annotation Description Scheme Standard (Chiariglione, 2006) go some way towards dealing with this problem, the use of a limited form of manual input cannot be fully avoided. The methods of extraction of the main kinds of M³O data are given in Table 2.

Table 2 Methods of data extraction for different mobile multimedia objects

<i>Data</i>	<i>Type of data</i>	<i>Method of extraction</i>
Keywords	Content data	Direct automatic extraction for text documents; For voice memos and video clips, indirect automatic extraction after conversion to text by speech recognition algorithms; For hand-written documents, automatic extraction via OCR
Musical key	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Visual objects and their relationships	Content data	Manual via MPEG-7 Structured annotation description scheme
Colour	Content data	Automatic using established image processing algorithms
Texture	Content data	Automatic using established image processing algorithms
Shape	Content data	Automatic using established image processing algorithms
Persons and Characters	Content data	Manual via MPEG-7 Structured annotation description scheme
Timing	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Note sequence	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Tempo	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Chord sequence	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Instrument	Content data	Usually manual for MP3; Automatic for MIDI via music composing software
Date/time	Metadata	Automatic
Originator	Metadata	Automatic for SMS, MMS; Mostly manual otherwise
Title/file name	Metadata	Automatic
Location	Metadata	Automatic via GPS
Artist	Metadata	Semi-automatic via ID3/CDDB
Composer	Metadata	Semi-automatic via ID3/CDDB
Duration	Metadata	Automatic
Genre	Metadata	Semi-automatic via ID3/CDDB
Size	Metadata	Automatic
Region name	Metadata	Semi-automatic via GIS
Salient features	Content data	Semi-automatic via GIS

4 Requirements of an M³O management system

In mobile multimedia object management, identification and retrieval of objects for purposes such as play, display, and propagation tend to be high. As there is limited scope for browsing or relevance feedback due to constraints in both time and display size, a high level of precision is required. Thus, a simple and precise query interface is essential; voice input and natural language are preferred, while the mechanical entering

of alphanumeric information (e.g., keyboard, pen input) should be kept to minimum. In addition, the system should support a good degree of browsing and fuzzy search. Consequently, strict database design rules such as normalisation should not be used since these are mostly geared to update operations and would introduce substantial inefficiency in relation to browsing. Likewise, conventional data management, concurrency and recovery procedures are unimportant, and designs that are geared to high transaction activity levels will also be inappropriate.

An M³O Management System must be lightweight in that it should not take up to much processing and storage overheads. Fully inverted indexes will not be feasible due their size and complexity. Although update to an M³O is mostly not required, the editing of an existing object to produce a new object such as SMS and MMS can occur from time to time. It is unlikely that the large-scale editing of text-based objects will occur on most mobile devices (except for laptop computers). Insertion/creation activities for M³O will be common due to the exchange and propagation of objects from other mobile devices. Table 3 shows the activity levels of the different data operations on M³O.

Table 3 Data operations for different mobile multimedia objects

<i>Data operation</i>	<i>Activity level</i>
Retrieval/Play/Display of object	Very high
Insertion/Reception of object	Medium to high
Creation of object	Medium
Deletion of object	Low to medium
Edit object	Low
Update of object	Very low
Navigation within object	Not required (except for maps)
Search within object	Not required (except for maps, large text-based documents, long video clips)

Since the first two data operations are the most common, a flexible M³O management system should be able to perform these efficiently as well as provide the ability to support a variety of retrieval paths using multiple criteria. In particular, with the first data operation, most current mobile devices are only able to support it at the most basic level, which is relatively rigid and generally unable to satisfy the range of functions and requirements of the discerning user.

5 Data architecture of the M³O management system

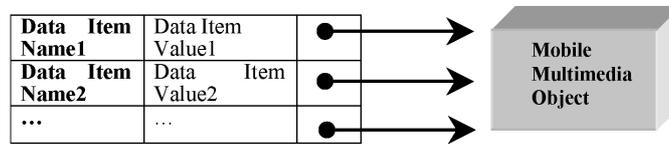
As indicated above, search is very important and is most commonly used. Since different types of objects have their own search requirements and characteristics, search indexes will need to be tailored for different categories of M³Os. These indexes will reveal the object ID, while the retrieval and display mechanisms will be undertaken by hardware. There are two levels of search: basic search and sophisticated search, each having their own data requirements to support the search. The basic level will be based primarily on metadata. The sophisticated level will be based mostly, but not exclusively, on content.

Basic search tends to involve retrieval criteria that are common and universal, which apply across most M³O_s. As indicated in the previous sections, these would include the file name/title, date/time, size/length and originator. To support content-based and high-precision search, these are often inadequate, and an additional secondary object – the M³O index object or M³I – which is linked to the M³O is required. Each M³O will have its own M³I for access purposes, and the M³I will possess the following properties.

- *diversity of data among different types of M³O_s*: for example, an image may require information on the visual objects and their relationships within the image, while a song may require information on the composer, solo instrument or lyrics
- *heterogeneity within the same type of M³O*: for example, in some songs, the key and the artist data are available, while in others, such data may be unavailable; sometimes, the M³I may be completely absent for some M³O_s.

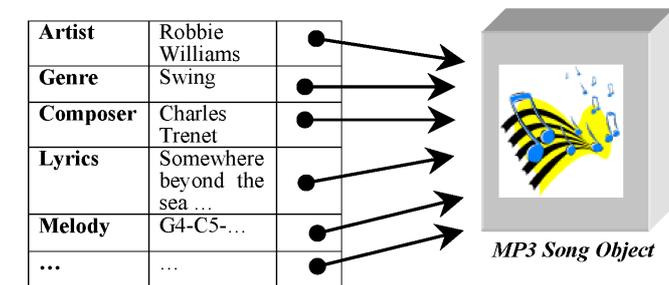
The design of a flexible and interoperable M³I structure must cater for such diversity, and for this reason, the standard fixed-length record normalised relational database will not be efficient both in terms of storage utilisation, and in terms of browsing flexibility. To facilitate exchange and interoperability, the index structure must be self-descriptive with standardised field names so that there is no ambiguity in the meaning of the data item, and where the same data items are involved, these can be correctly understood and processed across multiple devices. The M³I has the general structure shown in Figure 1.

Figure 1 Structure of M³I



Figures 2–4 give examples of common M³I_s. Note that such an index object is built over and above the more basic level of data relating to the object such as filename, size, etc., which is already available.

Figure 2 M³I and M³O for MP3 object



Mobile Multimedia Index Object M³I

Figure 3 M³I and M³O for MP4 object

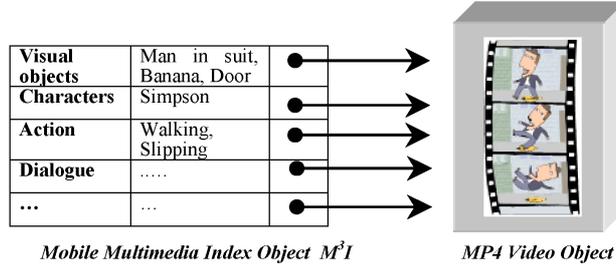
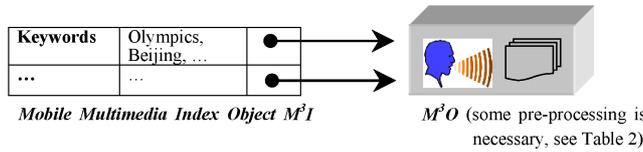
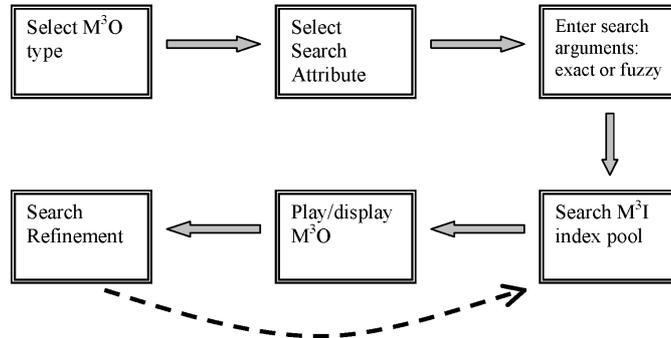


Figure 4 M³I and M³O for voice memo/SMS/hand-written objects



While an M³O is not searchable, its M³I is searchable. For efficient search and management, M³Is are normally categorised according to the M³O types, and M³Is belonging to the same type of M³O are clustered together. Typical access to an M³O by content is shown in Figure 5.

Figure 5 Accessing M³O by content via M³I

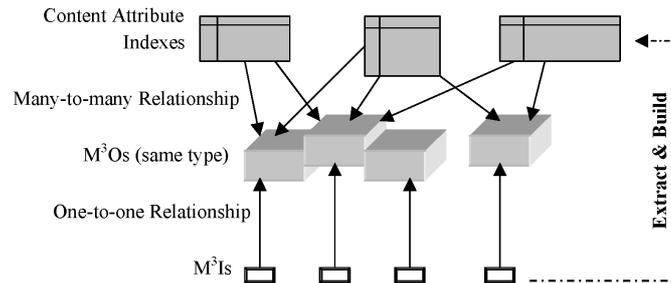


It is possible to retrieve more than one M³O meeting the search criteria in which case they can either be displayed sequentially or, preferably, their chief distinguishing attribute may be shown in a list for refining the search. For example, if ‘Artist’ is searched, then the different genre of songs from that artist may be listed for further selection. The dotted arrow in Figure 5 signifies a possible iterative loop executed multiple times. The present architecture provides flexibility in two ways.

By not adhering to a rigid structure, the addition or omission of a particular item will not cause any problems. For example, in Figure 3, the visual objects may be augmented by a further MPEG-7 description; alternatively, it may be assumed that an MPEG-7 description has been pre-processed to produce the visual objects, characters, and actions so that its inclusion is not necessary. Likewise, in Figure 2, the song lyrics may be augmented by a keyword list extracted from the lyrics to facilitate content search.

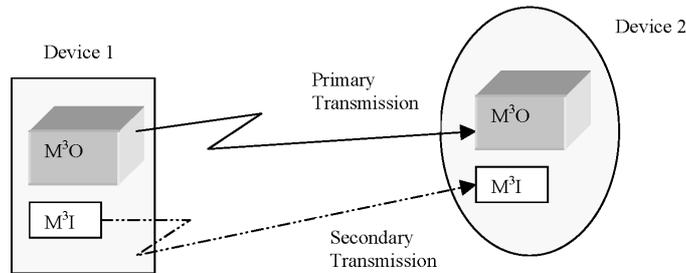
By including a separate pointer (i.e., Object ID) to the M³O for each data item, the index triplets (Data Item Name, Data Item Value, Object ID) need not be clustered together for the same object. In large-scale, high-efficiency search, a further storage structure clustering together like items for different objects (belonging to the same type) may be used, e.g., all the characters relating to different MP4 objects may be stored together as Content Attribute Indexes (see Figure 6), which are extracted and built from the M³Is.

Figure 6 Multiple content attribute indexes for the same object type



When an M³O is sent, its associated M³I will also be automatically obtained and transmitted in its entirety. This is shown in Figure 7, where a transmission of the M³O (primary transmission) will automatically trigger the transmission of the associated M³I (secondary transmission). The breaking down of the M³Is into individual components to form content attribute indexes, while raises the performance of individual systems, is unlikely to be supported by different systems in devices in a uniform way, and will not therefore be required for interoperability.

Figure 7 Primary transmission and secondary transmission



6 Summary and conclusions

The provision of only basic metadata for mobile multimedia objects represents a severe limitation in their retrieval and usage. Such a mode of object management is mostly inherited from traditional systems, which have been primarily geared to text-oriented object operations and is not able to satisfactorily cope with the diverse requirements of multimedia objects such as pictures, songs, and videos. For mobile devices, truly multimedia objects tend to constitute the dominant object type, while for conventional non-mobile systems, text-based objects such as documents, structured database tables,

spreadsheets have been dominant. Thus, while content search within a text-based object is relatively straightforward, content search of a song or video is far from being so. Computing systems have yet to fully adapt to the management and exchange of M³O's so that the full potential of these objects can be exploited.

In this paper, we have categorised the content requirements of different mobile multimedia objects, and described a flexible and portable data architecture, which is able to meet their exchange and usage needs. We have shown that for every M³O, there is an associated M³I attached to it, both of which are transmitted during an exchange. The M³I allows for a variety of content data to be incorporated, without adhering to a rigid structure. Such data architecture acknowledges and caters to the fact that there is significant diversity both between different multimedia object types as well as within the same type of objects. Depending on the size and complexity of the particular mobile device, mechanisms for performance improvement and optimisation can also be incorporated.

References

- Azzam, I., Leung, C.H.C. and Horwood, J. (2004a) 'Implicit concept-based image indexing and retrieval', *Proceedings of the 10th International Conference on Multi-media Modeling*, Brisbane, Australia, January, pp.354–359.
- Azzam, I.A., Charlapally, A.G., Leung, C.H.C. and Horwood, J.F. (2004b) 'Content-based image indexing and retrieval with XML representations', *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing*, Hong Kong, October, pp.181–185.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999) *Modern Information Retrieval*, Addison-Wesley, Boston, MA.
- Banavar, G., Becky, J., Gluzberg, E., Munson, J., Sussman, J. and Zukowski, D. (2000) 'An application model for pervasive computing', *Proceedings of the 6th ACM/IEEE International Conference on Mobile Computing and Networking*, Boston, pp.266–274.
- Bharat, R. and Minakakis, L. (2003) 'Evolution of mobile location-based services', *Communications of the ACM*, Vol. 46, No. 12, December, pp.61–65.
- Chiariglione (2006) <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
- Ghias, A., Logan, J. and Chamberlin, D. (2002) 'Query by humming: music retrieval in an audio database', in Jeffay, K. and Zhang, H.J. (Eds.): *Multimedia Computing and Networking*, Morgan Kaufmann, San Francisco, CA, pp.216–221.
- Gemmell, J. and Lueder, R. (2003) 'The mylifebits lifetime store', *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence*, Berkeley, November, pp.80–83.
- Graham, A., Garcia-Molina, H., Paepcke, A. and Winograd, T. (2002) 'Time as essence for photo browsing through personal digital libraries', *Proceedings ACM Joint Conference on Digital Libraries*, Portland, June, pp.326–335.
- Leung, C.H.C. and Sutanto, D. (1999) 'Multimedia data modeling and management for semantic content retrieval', in Furht, B. (Ed.): *Handbook of Multimedia Computing*, CRC Press, Boca Raton, pp.43–54.
- Leung, T.L.Y. (2007) *Query by Humming*, Project Report, Computer Science Department, University College London, University of London.
- Over, P., Leung, C.H.C., Ip, H. and Grubinger, M. (2004) 'Multimedia retrieval benchmarks', *IEEE Multimedia*, Vol. 11, No. 2, pp.80–84.
- Platt, C., Field, B. and Czerwinski, M. (2002) *PhotoTOC: Automatic Clustering for Browsing Personal Photographs*, Technical Report, Microsoft Research, February, Redmond, WA.

- Poon, E. and Chen, D. (2004) 'Dynamic multimedia templates for users of wireless pervasive computing devices', *Proceedings IEEE 6th International Symposium on Multimedia Software Engineering*, Miami, pp.110–113.
- Rodden, K. (2003) 'How do people manage their digital photographs?', *Proceedings of the ACM Conference on Human Factors in Computing Systems*, Fort Lauderdale, April, pp.409–416.
- Smith, J.R., Mohan, R. and Li, C. (1999) 'Scalable multimedia delivery for pervasive computing', *Proceedings of the 7th ACM International Conference on Multimedia*, Orlando, pp.131–140.
- Stevenson, K. and Leung, C.H.C. (2005) 'Comparative evaluation of web image search engines for multimedia applications', *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo*, The Netherlands.
- Toyama, K., Logan, R., Roseway, A. and Anandan, P. (2003) 'Geographic location tags on digital images', *Proceedings of the 11th ACM international Conference on Multimedia*, Berkeley, November, pp.156–166.
- Viola, P. and Jones, M. (2001) 'Robust real-time object detection', *Proceedings of the International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, Sampling*, Vancouver, July.
- Wilhelm, A., Takhteyev, Y., Sarvas, R., van House, N. and Davis, M. (2004) 'Photo annotation on a camera phone', *Proceedings ACM Computer Human Interaction*, Austria, pp.234–238.
- Zheng, Z.J. and Leung, C.H.C. (1996) 'Automatic indexing for rapid content-based retrieval', *Proc. IEEE International Workshop on Multi-media Database Management Systems*, IEEE Computer Society Press, Blue Mountain Lake, NY, August, pp.38–45.